# Yapay zekâ yalan söylediğinde sesi hiç titremiyor

Dr. Öğr. Üyesi Kerem Gencer
AKÜ Bilgisayar Mühendisliği Bölümü Yapay Zekâ Anabilim Dalı Başkanı

*Birkaç yıl öncesine kadar bir fotoğraf gördüğümüzde "demek ki olmuş" derdik. Bir ses kaydı dinlediğimizde o sesin sahibinin gerçekten o sözleri söylediğinden şüphe etmezdik. Bugün ise elimizdeki telefonla birkaç dakika içinde hiç yaşanmamış bir olayın fotoğrafını üretmek, hiç söylenmemiş bir cümleyi tanıdık bir sesle söyletmek mümkün hale geldi. Yapay zekâ hayatımıza yalnızca kolaylık getirmedi; gerçeklik algımızın temellerini de sessizce yeniden şekillendirmeye başladı.*

Bu değişimi anlamak için önce şu soruyu sormak gerekiyor. Yapay zekâ dediğimiz sistemler aslında ne yapıyor? Bugün konuştuğumuz sohbet robotları, görüntü üreten programlar ve öneri sistemleri, milyarlarca örnekten öğrendikleri istatistiksel örüntüleri kullanarak "en olası" cevabı üretiyor. Yani bu sistemler dünyayı bizim gibi yaşayarak, dokunarak, hata yapıp bedelini ödeyerek öğrenmiyor; insanların ürettiği devasa metin ve görüntü yığınlarından çıkardıkları kalıpları ustaca yeniden birleştiriyor. Sonuç çoğu zaman etkileyici, hatta insan elinden çıkmış kadar doğal görünüyor. İşte asıl mesele de tam burada başlıyor.

## İkna edici olmak ile doğru olmak aynı şey değil

Yapay zekâ sistemlerinin en dikkat çekici özelliği, yanlış bir bilgiyi de doğru bir bilgiyi anlattıkları aynı özgüvenle sunabilmeleri. Biz buna teknik olarak "halüsinasyon" diyoruz; sistem, hiç var olmayan bir kaynağı, hiç yaşanmamış bir olayı ya da yanlış bir tarihi son derece akıcı ve inandırıcı bir dille anlatabiliyor. Üstelik bunu kötü niyetle değil, çalışma prensibinin doğal bir sonucu olarak yapıyor. Sistem bir sonraki kelimeyi tahmin ederken "bu bilgi doğru mu" diye sormuyor, "bu cümle akıcı mı" diye soruyor.

Benim de araştırma alanım olan güvenilir yapay zekâ, tam olarak bu sorunla uğraşıyor. Bir yapay zekâ sisteminin yalnızca doğru cevap vermesi yetmiyor; ne kadar emin olduğunu da doğru ifade edebilmesi gerekiyor. Buna teknik dilde kalibrasyon diyoruz. Hava durumu tahminini düşünün. "Yarın yüzde yetmiş ihtimalle yağmur yağacak" diyen bir sistem, bu tür tahminlerinin gerçekten yüzde yetmişinde haklı çıkıyorsa güvenilirdir. Aynı ilke tıpta, hukukta, finansta kullanılan yapay zekâ sistemleri için hayati önem taşıyor. Bir hastalık riskini değerlendiren sistemin "eminim" dediğinde gerçekten emin olması, "emin değilim" diyebildiğinde ise kararı insana bırakması gerekiyor. Çalışmalarımızda gördüğümüz tablo şu ki, günümüz sistemleri çoğu zaman gerçekte olduklarından daha emin görünüyor. Bu da kullanıcıda yanlış bir güven duygusu yaratıyor.

## Gerçeklik algımız nasıl değişiyor?

Bu teknik meselenin toplumsal karşılığı çok daha geniş oluyor. Sahte görüntü ve sesler, yani kamuoyunda "deepfake" olarak bilinen içerikler, artık yalnızca eğlence amaçlı kullanılmıyor; telefon dolandırıcılığından, tanınmış hekimlerin ya da sanatçıların sesiyle hazırlanmış sahte ürün reklamlarına kadar uzanan bir yelpazede karşımıza çıkıyor. Ancak bence asıl tehlike sahte içeriğin kendisi değil. Asıl tehlike, her şeyin sahte olabileceğini bilen bir toplumun zamanla hiçbir şeye inanmamaya başlaması. Buna literatürde "yalancının kazancı" deniyor; gerçek bir görüntü ortaya çıktığında bile "bu yapay zekâ ürünüdür" deyip sıyrılmak mümkün hale geliyor. Yani sahtecilik yalnızca yalanı güçlendirmiyor, gerçeğin de değerini düşürüyor.

Bir diğer sessiz değişim de bilgiye ulaşma alışkanlıklarımızda yaşanıyor. Eskiden bir konuyu araştırırken farklı kaynakları karşılaştırır, çelişkileri görür, kendi yargımıza kendimiz varırdık. Şimdi ise tek bir sohbet penceresinden, tek bir sesle, derli toplu bir cevap alıyoruz. Bu büyük bir kolaylık, ama cevabın nereden geldiğini, hangi kaynaklara dayandığını ve nerede hata yapmış olabileceğini sorgulamayı unutursak, düşünme işini de yavaş yavaş devretmiş oluyoruz.

## Peki ne yapmalıyız?

Karamsar bir tablo çizmek istemem, çünkü gerçekte durum umutsuz değil. Yapay zekâ tıpta erken tanıyı, eğitimde kişiselleştirilmiş öğrenmeyi, bilimde yıllar sürecek keşifleri aylara sığdırmayı mümkün kılıyor. Sorun teknolojinin kendisinde değil, onunla kurduğumuz ilişkide. Bu ilişkiyi sağlıklı kurmak için birkaç basit ilke yeterli olacaktır.

Birincisi, yapay zekâdan gelen her bilgiyi bir uzman görüşü gibi değil, işini iyi bilen ama bazen kendinden fazla emin konuşan bir asistanın önerisi gibi karşılamak gerekiyor. Önemli kararlarda, tıpkı ikinci bir hekim görüşü alır gibi, bağımsız bir kaynaktan doğrulama yapmak hâlâ en sağlam yöntemdir. İkincisi, duygularımıza en çok dokunan içeriklere karşı en temkinli olmalıyız; çünkü manipülatif içerikler tam da öfkemizi, korkumuzu ya da hayranlığımızı hedef alarak yayılıyor. Bir görüntü ya da haber bizi anında paylaşmaya itiyorsa, önce bir nefes alıp kaynağını sorgulamak gerekiyor. Üçüncüsü, çocuklarımıza ve öğrencilerimize yapay zekâyı yasaklamak yerine, onu sorgulayarak kullanmayı öğretmeliyiz. Gelecek nesillerin en değerli becerisi bilgiye ulaşmak değil, ulaştığı bilginin güvenilirliğini tartabilmek olacaktır.

Biz araştırmacılar da kendi payımıza düşeni yapmaya çalışıyoruz. Yapay zekâ sistemlerinin belirsizliklerini dürüstçe ifade edebilmesi, kritik kararlarda insana danışmayı bilmesi ve verdiği cevapların hesabını verebilmesi için çalışıyoruz. Hedefimiz her şeyi bilen bir makine değil; ne bildiğini ve ne bilmediğini ayırt edebilen, bu ayrımı da bize açıkça söyleyebilen sistemler geliştirmek.

Sonuç olarak yapay zekâ ne bir mucize ne de bir felakettir. O, insanlığın ürettiği en güçlü araçlardan biri ve her güçlü araç gibi nasıl kullanılacağı onu kullanan kişiye göre değişiyor. Gerçeklik algımızı koruyacak olan şey teknolojiden kaçmak değil, onu tanımaktır. Makinelerin ne kadar ikna edici olabildiğini bilen bir toplum, kolay kolay yanıltılamaz. Bu yüzden yapay zekâ çağında en insani görevimiz belki de şu; merakımızı kaybetmeden şüpheciliğimizi, teknolojiye açıklığımızı kaybetmeden eleştirel aklımızı korumak olacaktır.